NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

CONCEPT OF OPERATIONS (CONOPS) FOR FOREIGN LANGUAGE AND SPEECH TRANSLATION TECHNOLOGIES IN A COALITION MILITARY ENVIRONMENT

by

Susan LaVonne Marshall

March 2005

Thesis Advisor: James F. Ehlert
Second Reader: Steven J. Iatrou

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE March 2005	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE: Title (Mix case letters) Concept Of Operations (Conops) For Foreign Language And Speech Translation Technologies In A Coalition Military Environment			**5. FUNDING NUMBERS**
6. AUTHOR(S) Susan LaVonne Marshall			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited	12b. DISTRIBUTION CODE A

13. ABSTRACT (maximum 200 words)

   This thesis presents Concept of Operations (CONOPS) for two specific automated language translation (ALT) devices, the P2 Phraselator and the Voice Response Translator (VRT). The CONOPS for each device are written as Appendix A and Appendix B respectively. The body of the thesis presents a broad introduction to the present state of ALT technology for the reader who is new to the general subject. It pursues this goal by introducing the human language translation problem followed by nine characteristic descriptors of ALT technology devices to provide a basic comparison framework of existing technologies. The premise is that ALT technology is presently in a state where it is tackled incrementally with various approaches. Two tables are provided that illustrate six commercially available devices using the descriptors. A scenario is then described in which the author observed the two subject ALT devices (depicted in the CONOPS in the Appendices) being employed within an international military exercise. Some unique human observations associated with the use of these devices in the exercise are discussed. A summary is provided of the Department of Defense (DOD) process that is exploring ALT technology devices, specifically the Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration ACTD

14. SUBJECT TERMS Human Systems, Human Language Translation, Automated Human Language Translation, Automated Language Translation, ALT, Machine Language Translation, Voice Response Translator, VRT, P2 Phraselator, Concept Of Operations, CONOPS, Language And Speech Exploitation Resources, LASER, Speech-To-Speech Translation, Phrase Based Translation	15. NUMBER OF PAGES 81
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

i

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

CONCEPT OF OPERATIONS (CONOPS) FOR FOREIGN LANGUAGE AND SPEECH TRANSLATION TECHNOLOGIES IN A COALITION MILITARY ENVIRONMENT

Susan LaVonne Marshall
Lieutenant Commander, United States Navy
B.S., Georgetown University, 1989

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
March 2005

Author: Susan LaVonne Marshall

Approved by: James F Ehlert
                 Thesis Advisor

                 Steve Iatrou
                 Second Reader

                 Dan C. Boger
                 Chairman, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis presents Concept of Operations (CONOPS) for two specific automated language translation (ALT) devices, the P2 Phraselator and the Voice Response Translator (VRT). The CONOPS for each device are written as Appendix A and Appendix B respectively. The body of the thesis presents a broad introduction to the present state of ALT technology for the reader who is new to the general subject. It pursues this goal by introducing the human language translation problem followed by nine characteristic descriptors of ALT technology devices to provide a basic comparison framework of existing technologies. The premise is that ALT technology is presently in a state where it is tackled incrementally with various approaches. Two tables are provided that illustrate six commercially available devices using the descriptors. A scenario is then described in which the author observed the two subject ALT devices (depicted in the CONOPS in the Appendices) being employed within an international military exercise. Some unique human observations associated with the use of these devices in the exercise are discussed. A summary is provided of the Department of Defense (DOD) process that is exploring ALT technology devices, specifically the Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration ACTD.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION AND OVERVIEW

A. PURPOSE

As the title of this document suggests, its primary purpose is to provide Concept of Operations (CONOPS) for use of automated language translation (ALT) technologies in a coalition military environment. To achieve this goal, two specific ALT devices were chosen by the author and a CONOPS for each one has been written as Appendix A and Appendix B to this document. Although it is unorthodox to answer the "thesis" question in the appendices, rather than in the body, it works well in this instance for the following reasons. First, the sponsor of this thesis specifically requested CONOPS for these two devices and for supporting documents to be self contained for ease of further routing within the acquisition process. Second, the format of Appendix A and Appendix B is consistent with other CONOPS for other technologies being routed through the same type of acquisition process. That format differs from the NPS thesis format so breaking the CONOPS out as Appendices satisfies both format requirements.

Given that the thesis question is answered in the Appendices, the logical next question is "what is the body of the thesis about"? In short, it is a broad introduction to the overall present state of ALT technology for the reader who is new to the general subject. It pursues this goal by introducing the human language translation problem in the next section. Then in Chapter II, nine characteristic descriptors of ALT technology devices are offered to provide a basic comparison framework of existing technologies. The premise is that ALT technology is presently in a state where it is tackled incrementally with various approaches. Chapter III goes on to describe a scenario in which the author observed the two subject ALT devices (depicted in the CONOPS in the Appendices) being employed within an international military exercise. It explores some unique human observations associated with the use of these devices in a face-to-face scenario with a foreign national person. Chapter IV provides a summary of the Department of Defense (DOD) process that is exploring ALT technology devices, specifically the Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD). The Program Manager for the LASER ACTD is the sponsor of this thesis. Overall the body of this thesis is a broad introduction

to those unfamiliar with the subject and attempts to present it at a level that will encourage familiarity without delving too deeply in a subject that can quickly get very complex.

B.   DISCUSSION

The notion that human language translation can be accomplished by technology and machines is an appealing one. Star Trek fans are familiar with the "Universal Translator". It allowed Captain Kirk and his crew to communicate with inter-planetary aliens in real time. The reality of 21st century Earth, though, is that human machine language translation is still a tremendous challenge for technology. There does not exist yet a "Star Trek Universal Translator", this capability is probably decades away. In the meantime though, the process of pursuing real time ALT technologies has not presented itself in a neat linear scale but rather as an abundance of different devices representing different approaches and methods.

Before introducing the vocabulary it is essential to understand the problem. Anyone who has ever traveled to a foreign country and felt the pain of not being able to communicate with the local populace already has a sense of it. On a national scale, there are tremendous political and military issues associated with human language translation. Both the DOD and the Intelligence Communities (IC) need human language processing capabilities in a wide range of languages—for use with both speech and text—to support coalition/joint task force headquarters and tactical or routine field operations. Whether handling tactical intelligence or handling foreign national personnel seeking coalition medical assistance, the need for human language translation exceeds the availability of linguists.[1] ALT Technologies can and should increasingly fill this gap, especially as the technologies become more capable.

The DoD Operational Community deploys Joint forces worldwide. Most often, units deploy with insufficient numbers of qualified specialists in languages needed to support existing mission requirements. Foreign language support in the continental United States via reach-back is equally lacking. Joint forces are increasingly becoming coalition forces and there are many exercises being conducted annually with coalition

---

1 Office of the Secretary of Defense, *Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD) Management Plan*, November 2003, 5.

partners. Language capability is essential in force protection for deployed forces, humanitarian, and peacekeeping operations as well as tactical and operational intelligence operations.

The IC is faced with a vast increase in collection capabilities and availability of open source information in widely diverse languages. Projected increases in baseline collection capabilities will further exacerbate the imbalance between what can be collected and what can be analyzed, especially by front line intelligence units. There needs to be some help in sorting through the mass of collection, i.e., some sort of triage system to more quickly translate, identify and sort out relevant material. Foreign language capable personnel, augmented by language translation related technology, could be fundamental to the collection, processing, and exploitation of these foreign language materials and sources.

THIS PAGE INTENTIONALLY LEFT BLANK

II. TYPES OF LANGUAGE TRANSLATION TECHNOLOGIES

Comparing and categorizing contemporary language translation technologies requires the reader to understand specialized terminology. This chapter offers descriptors, grouped as "primary" and "secondary". This list of descriptors is not intended to be a complete dissection but rather a functional baseline for discussing contemporary ALT devices. The primary descriptors, of which there are three, represent the highest order grouping of devices. They are considered primary because they have a significant effect on what the device may look like, what missions it is used in, and how much time lag it experiences. Any conversation about a particular device will almost always start with a sentence that identifies these three descriptors. For instance, "the Voice Response Translator is a *speech-to-speech*, *one-way*, *phrased-based device*". The secondary descriptors provide useful comparative information at a finer level of granularity. As the technologies mature and the devices become more capable, some of these descriptors will likely begin to blend together. The ultimate eventual device, the notional Star Trek "Universal Translator", probably would not need any of these descriptors.

It is worth noting that none of these descriptors address quantitative or qualitative performance measurements. This is deliberate because it is difficult to measure and identify performance metrics across dissimilarly constructed devices.

A. PRIMARY DESCRIPTORS

1. "Speech-to-Speech" or "Text-to Text"

Speech-to-speech is translation that is typically initiated by a voice speaking in the source language into a microphone input or selecting a written input from a screen and the resulting target language translation is produced audibly via an audio device such as a speaker.

Text-to-text is translation that is initiated and produced via text, such as on a computer keyboard and screen.

A typical speech-to-speech device is usually a stand-alone device with at least a microphone and a speaker. Sometimes it is mounted in a Personal Data Assistant (PDA)

type device and sometimes it is mounted on a laptop computer. A text-to-text device is usually on a computer with a keyboard and monitor screen showing the translation prose in both the source and the target language. In some cases there are several computers connected in a network to facilitate an instant message type "chat" environment. Text-to-text may use Optical Character Recognition (OCR) to scan written foreign language documents as well.

Sometimes a device can do part of both speech-to-speech and text-to-text, such as in the case where the user speaks an input and the device responds by presenting more than one written option to select from. The user then selects the most appropriate response and the device broadcasts the translation.

2. "One-Way", "One-and-a-Half-Way", or "Two-Way"

One-way translation is translation from a source language into a target language.

One-and-a-half-way translation is translation from a source language to a target language and from the target language back to the source language if the response falls within a set of expected responses. For instance, if a medical person asks a patient "where does it hurt?", the device will translate the reply as long as it is something like "my leg hurts". It will not translate a reply such as "it is raining" because this is not in the realm of expected responses to the question of "where does it hurt?"[2]

Two-way translation is translation from a source language into a target language and from a target language back into the source language. [3]

A one-way translator obviously has less utility than a two way translator. Given that there are many simple situations where one way translation is enough, a one-way translator affords a less technically challenging and expedient solution. Two-way translation significantly increases the technological challenge. An example of a simple one way scenario would be connecting an ALT device to a loudspeaker on a ship and warning approaching foreign boats to turn away or face being fired upon.

---

[2] Breault, Chris of the US Marine Forces Pacific Experimentation Center. Private conversations 13 Aug 04 through 18 Oct 04.

[3] Department of Defense. *"Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration Community Assistance Response Exercise (CARE) 2004 Assessment Execution Document (AED)"*. May 2004, 3.

3. "Phrase-Based" or "Free-Flowing"

Phrase-based translation relies on speech recognition software to identify specific speech input in the source language and match it to a pre-recorded phrase in a target language. The input can be the phrase itself (e.g., "Put your hands in the air") or a simple command that stands for the phrase (e.g., the command "Warning 1" would be programmed as "Put your hands in the air"). The same concept of matching phrases also exists in text-to-text translation and is sometimes called "word/phrase based translation".

Free-flowing translation uses computer processing to translate any words or sets of words from a source language input into another language with equivalent meaning.[4]

A phrase based device is the easiest to create from a technical standpoint. In a very basic sense, it is nothing more than matching pre-recorded sound bites. This is analogous to recording phrases in a tape recorder and then playing them back. The complexity lies mostly in the speech recognition capability of the device to recognize the actual phrase in the source language and then ensure it matches it with the correct translated phrase and broadcasts it accordingly. There does exist some technology that can recognize phrases imbedded within sentences, as opposed to matching only exact phrases. This "filtering" of phrases is still basically "phrase based" in concept but more technically complex.

Free flowing translation is usually accomplished by employing a machine translation (MT) engine used in conjunction with a word/phrase based Translation Memory (TM) and possibly some specialized domain specific dictionaries. The MT engine performs algorithmic translation (via one of about three existing approaches beyond the scope of this document) while the TM is populated manually by the user for commonly used words, phrases or acronyms particular to the user. For instance, the military uses many unique phrases and acronyms that repeat frequently. The MT engine can sometimes be programmed to use phrases from the TM based on minimum percentage search matches.

---

4 Air Force Operational Test and Evaluation Center (AFOTEC), Detachment 1. *"Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD) Community Assistance Response Exercise (CARE) 2004 Limited Military Utility Assessment (LMUA) Report."* July 2004, 5.

A phrase based device also typically experiences less time lag than a free-flowing device. Because a free flowing device has to algorithmically process all inputs, it simply needs more time to sort through the immense possibilities. Consider how the structure of human speech varies from language to language. In the German language, for instance, the verb is usually at the end of the sentence, so the machine translator has to grasp the content of the sentence and then reconstruct it. In the French language there is no word for "wife", the typical expression is simply "woman". The free flowing translator thus has to determine the context of the use of the word to determine if it should be "wife" or "woman". There is no magic number for how long it takes a machine translation engine to translate a phrase but in a recent technology "users" conference in San Diego, the author observed that the free-flowing translators had noticeable time lag from the input to the output, sometimes on the order of several seconds.

B. SECONDARY DESCRIPTORS

The secondary descriptors for describing a particular ALT device are more granular. Like the primary descriptors, they help to categorize ALT devices.

1. "Supported Domains"

Supported domains is a general reference to topics and sub-topics of use for the device. Some common high level domains include "medical" and "force protection" but may also include lower level component domains such as "medical triage", "medical processing", "refugee processing", "missing persons", "travel", "checkpoint", "maritime interdiction", and "DUI". This is by no means a complete list but rather a concept of grouping.

2. "Supported Languages", "Source Language" and "Target Language"

Supported languages are all of the languages included in the device.

Source language is the language of the device user, in most cases English.

Target Language is the language being translated to. Many devices have more than one target language.

3. "Speaker Dependent" or "Speaker Independent"

Speaker-dependent devices must be programmed to recognize the speech patterns of specific users. Such devices can be used effectively with only those individuals who have pre-recorded their voices to the device.

Speaker-independent devices can be used without being programmed to recognize the unique speech patterns of a specific user's voice.[5]

As the name implies, speaker dependent or speaker independent applies only to speech-to-speech devices and not to text-to-text devices.

4.   "Stand-alone" or "Network Based"

Stand-alone is a device that can be carried and used entirely by itself. This is normally in some form like a Personal Data Assistant (PDA), a smaller vest mounted device, or a laptop computer. Speech-to-speech devices are typically stand-alone devices because they must be highly mobile.

Network based is a device that relies on network of computers to execute its full resources.

5.   "Operating System"

Operating System refers to its computer operating system such as Windows, Linux, or proprietary code.

6.   "Technology Readiness Level (TRL)"

Technology Readiness Level is a scale from 1 to 9 that roughly describes the maturity of the system. This scale was created specifically for the LASER ACTD (see Chapter IV) and provides a rough indication of its usability. The TRL's are subjective so two different people may assign a different TRL for one particular device but they would most likely be close. Table 1 describes the nine TRL's.

TRL's are worth presenting in this venue because they avoid the difficulty of evaluating these devices quantitatively but still provide some sort of a useful opinion on their utility. Given that there are many variables to the question of "how well does it (the ALT device) work?", the TRL's bypass this question by focusing on "how ready is it - given what (type descriptors) it is?"[6]

Formal quantitative or qualitative evaluations of one single device require a large amount of resources due to the large number of variables and even then many of the conclusions would still be subjective. An excellent illustration exists in the question of

---

5 Ibid., 6.

6 Breault, Chris of the US Marine Forces Pacific Experimentation Center. . Private conversations 13 Aug 04 through 18 Oct 04.

"what percentage of translations are accurate?" The question implies a numerical response but there are two problems; what constitutes an "accurate translation" and what would be the point, given the type of device? On the first issue, five different linguists may not agree on one translation. On the second issue, how would one define accuracy of translation for a phrase based device versus for a free flowing device? The same subjective linguist opinion applies but less for pre-recorded phrases in phrase based devices. The linguists recording the phrases can take all the time they want to get it right before the device ever gets near a target subject. In free-flowing devices, where time as more the essence, a percent-accurate would be more useful but is again, subject to the opinion of the linguists.

Another illustration exists in the question "how long does it take?" The issue becomes what is the context of the situation, how long was the input, and what is the type of device? Opinions on performance of ALT devices are therefore subjective and very much dependent on what type of ALT device is being evaluated and what they are intended to do. For this reason, the descriptors in this chapter are limited to categorization-type rather than performance-type.

Table 1.   Technology Readiness Level Description. (From: The LASER ACTD Management Plan)

Technology Readiness Level	DESCRIPTION
1	Basic principles observed and reported. Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples might include paper studies of technology's basic properties.
2	Technology concept and/or application formulated. Invention begins. Once basic principles are observed, practical applications can be invented. The application is speculative and there is no proof or detailed analysis to support the assumption. Examples are still limited to paper studies.
3	Analytical and experimental critical function and/or characteristic proof of concept. Active research and development is initiated. This includes analytical studies and laboratory studies to physically validate analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.
4	Component and/or breadboard. Validation in laboratory environment. Basic technological components are integrated to establish that the pieces will work

Technology Readiness Level	DESCRIPTION
	together. This is relatively "low fidelity" compared to the eventual system. Examples include integration of "ad hoc" hardware in a laboratory.
5	Component and/or breadboard validation in relevant environment. Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so that the technology can be tested in a simulated environment. Examples include "high fidelity" laboratory integration of components.
6	System/subsystem model or prototype demonstration in a relevant environment. Representative model or prototype system, which is well beyond the breadboard tested for technology readiness level (TRL) 5, is tested in a relevant environment. Represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high fidelity laboratory environment or in a simulated operational environment.
7	System prototype demonstration in an operational environment. Prototype near or at planned operational system. Represents a major step up from TRL 6, requiring the demonstration of an actual system prototype in an operational environment, such as in an aircraft, vehicle or space. Examples include testing the prototype in a test bed aircraft.
8	Actual system completed and "flight qualified" through test and demonstration. Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications.
9	Actual system completed and "flight qualified" through test and demonstration. Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications.

C.   SAMPLE DEVICES

Tables 2 and 3 offer specific examples using the terminology described in this chapter. Table 2 contains speech-to-speech devices and Table 3 contains text-to-text devices. They are separate tables in this manner because several of the secondary descriptors only apply to either a speech-to-speech device or to a text-to-text device. The tables are not intended to describe each device in depth but rather to present a broad comparative overview to illustrate the descriptors discussed above. Each of these devices could arguably be the subject of its own thesis if one chose to examine it in depth.

Additionally, it is worth noting that hundreds of devices are commercially available, these six are merely the most readily accessible to the author.[7,8,9,10,11,12,13]

Table 2.   Speech-To-Speech Automated Language Translation Device Samples

Product Name	Voice Response Translator (VRT)	P2 Phraselator	S-Minds
Manufacturer	Integrated Wave Technologies	VOXTEC	Sehda Inc
One-Way, One and-a-Half Way or Two-Way	One Way	One Way	One-and-a-Half Way
Phrase Based or Free Flowing	Phrase Based	Phrase Based	Phrase Based with more than one choice for same phrase and close-enough-type matching
Supported Domains	Force Protection, Medical, Logistics, Law Enforcement, Maritime Interdiction Operation (MIO)	32 "Phrase Modules" available containing at a minimum: Force Protection, Medical, Disaster Relief, Maritime Interdiction Operation (MIO)	Up to six domains available depending on language: Medical, Ship Boarding, Maps/Directions, Force Protection, Refugee Processing
Supported Languages	30 languages including Korean, Thai, Iraqi, Spanish	35 languages including Arabic, Spanish, French and	Korean, Japanese, Spanish, Serb-Croatian, Arabic-

7 Hall, John of Integrated Wave Technologies. Private telephone conversations 29 Nov 04 through 7 Dec 04. Monterey, CA.

8 Sehda Inc. Solutions S-Minds web-page. http://www.sehda.com/solutions.htm. (Accessed 21 Feb 05).

9 Speechgear Compadre Expres web-page. http://www.speechgear.com/compadre.aspx (accessed 25 Oct 04)

10 Hall, John of VOXTEC. Private telephone conversations 18 Feb 04 through 7 Mar 04. Monterey, CA

11 LeBlanc, Ray of MITRE Corporation. Private telephone conversations 28 Feb 05 through 3 Mar 05. Monterey, CA

12 Phraselator Model P2 web-page. http://www.phraselator.com/products/prod_p2.aspx (accessed 27 Feb 05)

13 Ehsani, Farzad of Sehda Inc. Private telephone conversation 28 Feb 05. Monterey, CA.

Product Name	Voice Response Translator (VRT)	P2 Phraselator	S-Minds
	and Pashto	Korean	Iraqi
Speaker Dependent or Speaker Independent	Speaker Dependent	Speaker Independent	Speaker Independent
Stand Alone or Network Based	Stand Alone, mountable on a vest	Stand Alone, PDA style	Stand Alone, on a laptop
Operating System	Proprietary Code	WinCE.NET 4.2	Windows
Technology Readiness Level (from Table 1)	7	7	7

Table 3.   Text-To-Text Automated Language Translation Device Samples

	FALCON	Trans-Instant Messaging (TrIM)	Expres
Manufacturer	Integrated products under the Army Research Laboratory (ARL)	Integrated products under MITRE	Speech Gear
One-Way, One and-a-Half Way or Two-Way	Can be One Way or Two-Way depending on the language and which Machine Translation engine is supporting it.	Two Way	Two-Way
Free Flowing or Word/Phrase Based	Free Flowing with Word/Phrase Based Translation Memory and dictionaries	Free Flowing with Word/Phrase Based Translation Memory and dictionaries	Free Flowing with Word/Phrase Based Translation Memory and dictionaries
Supported Domains	Unlimited, determined by how well the TM is populated and which dictionaries are tied in	Unlimited, determined by how well the TM is populated and which dictionaries are tied in	Unlimited, determined by how well the TM is populated and which dictionaries are tied in
Supported Languages	Chinese, Japanese, Korean, Swahili, Pashto, Tagalog	Korean	Korean, Thai
Stand Alone or Network Based	Stand Alone or networked on a desktop or laptop	Network Based instant messaging "chat" on desktops	Stand Alone or networked on a desktop or laptop

	FALCON	Trans-Instant Messaging (TrIM)	Expres
	depending on where the MT engine is located.	and laptops.	depending on where the MT engine is located.
Operating System	Windows	The server is typically LINUX based. The network it connects into can be Windows	Windows
Technology Readiness Level (from Table 1)	7	7	7

D.   SUMMARY

This chapter has attempted to provide the reader with basic terminology and a framework for categorizing and discussing current ALT technology devices.   Three primary and six secondary descriptors were offered along with two tables illustrating the use of these descriptors with respect to a few actual devices currently on the market.

III. CURRENT HUMAN ISSUES WITH ALT DEVICES

A. INTRODUCTION

Contemporary ALT technologies do not function ubiquitously and in real time – nor are they close to doing so. The ideal Star Trek "Universal Translator" is still just a notion. In the meantime though, there exist many different devices representing different approaches and methods. A suitable analogy to describe the current state of automated language translation exists with human flight. Human beings cannot fly by themselves but they can fly with the assistance of many different types of devices, for instance a helicopter or a hang glider. Each device requires some learning and skill building until eventually the human being can exploit its full capability. The physical characteristics of the flight controls, and the approach to flying with a helicopter is different than flying with a hang glider. In fact it is hard to say they have much in common except that they both help humans fly. Current ALT technologies are similar in that they are very diverse in appearance and method but they can help humans communicate to each other in a foreign language. Like flying, this communication has limitations that must be understood by skill building and practice to achieve full potential. The full potential of present day ALT devices is not unlimited, but many possess a significant amount of utility provided the training is accomplished and the limits are well understood.

B. FIELD OBSERVATIONS

During a major South Korean – American military exercise in South Korea in August 2004, several agencies and individuals associated with the LASER ACTD (discussed in Chapter IV), were present - performing formal and informal evaluations and demonstrations of five types of automated language translation technologies. Two of these devices, the P2 Phraselator and the Voice Response Translator (VRT) were demonstrated and evaluated informally with the author of this thesis present and observing with the intent of writing military CONOPS for the devices. The P2 Phraselator and the VRT are each explained in extensive detail in their individual CONOPS, which are Appendix A and Appendix B of this thesis respectively. For purposes of the discussion in this chapter, the reader should know at a minimum that both

devices are *speech-to-speech, one-way, phrased-based* devices as explained in the definitions framework of Chapter II.

C.   THE SCENARIO

In the exercise, seven US Marines and six non-English speaking South Korean Marines were brought together to attempt using the P2 Phraselator and the VRT. The seven US Marines were Military Police ranging in rank from E-3 to E-6. They were provided with the devices and the associated instruction manuals on the first day. A LASER ACTD (see chapter IV) representative provided about one hour of verbal and visual instruction to the group and left the devices with them overnight. The Marines were encouraged to look up and become familiar (on their own) with the phrase lists and to specifically pick out those they would use in a gate-guard type scenario. They were informed that they would be asked to role play a gate-guard scenario the next day with the South Korean Marines.

The informal field demonstration/assessment was constructed around a gate guard scenario. The US Marines were instructed to role play as a gate guard to a US coalition compound while the South Korean Marines were told to approach the US Marine gate guard and seek entry to the compound. With the help of a linguist, each South Korean Marine was also given a role to play which included a basic set of instructions for who he was and whether or not he had an appointment and a weapon in his possession. Each South Korean Marine in turn then approached the US Marine guard and attempted entry into the compound. The US Marines had been instructed to allow entry only to those people with proper ID and an appointment. Additionally, personal weapons were to be confiscated and every person entering needed to be searched. The result in the case of all seven US Marines was that none of them were able to execute each scenario fully and correctly with the ALT device. Sometimes they forgot to verify an appointment, sometimes, they forgot to ask if the person was carrying a weapon, and sometimes they forgot to search the subject. It was as though the extra effort of employing the device made doing their basic job more difficult. It was also observed that US military personnel were quickly frustrated by the ALT devices and in some cases they "froze up" in the scenarios requiring prompting from observers.

Given that the devices have no formal classroom training structure in place beyond the enclosed instruction manual, it could be said that these US Marines received "extra training" by virtue of the one-hour session the day before with the LASER ACTD representative. It became apparent in the scenarios that a lot more familiarity and practice-type training was needed beyond just how to turn on the device and look up phrases.

D.   FINDINGS

1.   Expectation

The first human issue that created a barrier to using ALT devices in the above scenarios could be best described as "expectation". It was difficult for the participants to identify this point exactly so an analogy may help. In order to fly, human beings expect to have to use a device to assist them – for instance a helicopter or a glider. For human communication though, there is a very basic expectation of being able to communicate "as we are". People readily accept that humans need a technology device to help them fly but they do not readily accept that they need a technology device to help them communicate. After all, humans communicate in their native language all of the time and human linguists translate all of the time without technology. The important point is that current automated language translation technology is not mature enough that humans can expect it to behave like the Star Trek "Universal Translator" and there are never likely to be enough linguists.

Human beings communicate on many levels all of the time. They communicate with spoken and written language every day, plus with their body language. This is so integral to human existence that it hardly seems conscious, whereas flying is not integral to human existence and humans therefore accept more readily that they need a technology device assist. So the challenge for human beings is to accept that they need human language translation technology and to accept that it has limitations in its current state that will cause humans to have to spend some time learning these limits and practicing. In the South Korean exercise scenarios described above, the US Marine users clearly indicated they would prefer to have a linguist and although offered the opportunity for extra training with ALT devices, they declined. They did, however, indicate they could see the utility of the devices and thought they could be useful with

17

more practice and training. This is basically like saying "yes I see the utility but I do not want to do it".

2. Social Acceptability or Comfort

The second human issue that creates a barrier to using ALT devices is "social acceptability or comfort". It is not difficult to appreciate how useful it would be if everyone could communicate with anyone from any culture at any time. The reality, however, of approaching a foreign national person with a machine language translation device is that it is more confusing and intimidating than one would imagine. In the South Korean exercise scenarios described above, it was observed that the foreign national subject's initial reaction to an ALT speaking South Korean was simply confusion. The initial message played by the ALT device user was "this is a machine language translation device that speaks pre-recorded phrases from my language to your language, please nod your head yes if you understand so far". The initial response by all six South Koreans was confusion, which looked like a blank stare of disbelief. The ALT device user would then repeat the same phrase at which time the subject would visibly more focus their attention on the user and usually respond with an appropriate affirmative nod. It was as though the shock of seeing an obviously American person talking in Korean with a machine was too much too absorb on the first presentation.

After the initial shock wore off, though, there were still elements of body language by both the user and the subject indicating mutual discomfort. For instance there was a distinct lack of eye contact when executing the gate guard scenarios between the US Marines and South Korean Marine role playing subjects. This occurred even though it was pointed out to the US Marines that they should never relinquish eye contact in an actual gate guard situation. Taking one's eyes off of the subject is to relinquish control of the situation. Being uncomfortable, though, was apparently enough to induce this.

3. Socio-Cultural Differences

The third human issue could be described as "socio-cultural differences". This relates to the previous point about social acceptance and discomfort. There are cultural elements of communication that go beyond spoken or written words. Body language and gestures mean different things in different cultures. For instance, in Iraq, the gesture for

"no" is one upward nod of the head. This would appear to most Americans to look like "yes" or "go away". In Thailand, the gesture to beckon someone toward you is to turn the palm of the hand downward and repetitively curl the fingers inward – which is opposite the American gesture where the palm of the hand is upward. Additionally, there are body gestures that are offensive in some cultures and not in others. For instance in Arab cultures in general, it is considered rude to reach out with your left hand or to show the bottom of your foot. In other cultures, sustained eye contact is considered rude and that rule may vary depending on which sex is being addressed. To avoid a mistake in these instances, the ALT device user would need some definitive cultural training about how to say "yes" and "no" in the target language and what hand gestures are used to signal "come here" or "Okay". Any advantage gained by the use of an ALT device could quickly be lost by mistaking the visual response.

E.   SUMMARY/CONCLUSION

While it could be argued that humans are reluctant to accept any change and any new technology, the human issues described above were particular to the use of ALT devices. These issues are not obvious until observing someone trying to actually use an ALT device with a foreign national person, such as described in the military exercise in South Korea. ALT technology vendors and perspective users should be aware of these subtleties prior to selling and purchasing these devices. The devices do have utility but they will not help anyone if they remain in the box. Thorough understanding of the limits, human and technical, combined with the right kind of training, will ensure that users actually employ the devices.

The three human issues discussed above are mostly applicable to situations where the user is face-to-face with a foreign national person, such as when using a speech-to-speech device. In the realm of text-to-text devices, the same issues of social acceptance and discomfort may not exist since the user is basically interacting with a computer terminal and not a person. The challenges in text-to-text are likely more in the technical realm of developing more accurate and efficient Machine Translation engines, plus incorporating Optical Character Recognition technology for foreign language written material. A further discussion of the technical issues is beyond the scope of this thesis.

The next chapter shifts away from the specifics of employing an ALT device to provide a summary of the Department of Defense (DOD) process that is exploring ALT technology devices, specifically the Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD). The Program Manager for the LASER ACTD is the sponsor of this thesis.

IV.   OBJECTIVES AND APPROACH OF THE LASER ACTD

The Advanced Concept Technology Demonstration (ACTD) program was initiated in 1994 and is run under the Office of the Secretary of Defense (OSD). The purpose of an ACTD is to emphasize the assessment and integration of commercial or government technologies (as opposed to full blown research and development) to expedite the transition of maturing technologies from the developers to the users. An ACTD assembles its target technologies into an operationally useable form and inserts it into the operational environment to demonstrate new or improved military capability and utility. ACTD's demonstrate the use of such technologies to address critical military needs and are established based on response to user needs, maturity of technologies, and potential effectiveness of the technologies.

ACTD's are not themselves acquisition programs, but are designed to provide a residual, usable capability upon completion, and/or transition into acquisition programs. At the conclusion of an ACTD, there are three potential outcomes that the user sponsor may recommend:

- Acquisition and fielding of the residual capability that remains at the completion of the demonstration phase of the ACTD to provide an interim and limited operational capability

- Fielding of the residual capability without acquiring additional units if the user's need is fully satisfied

- Terminating the project or returning it to the technology base [14]

The Language and Speech Exploitation Resources (LASER) ACTD was initiated in Fiscal Year (FY) 02 under a three year program of demonstrations and a two year phase for transition of deliverables. LASER's objective is to demonstrate automated language technology devices, concepts and architecture paths to reduce human language barriers experienced by the DOD Operational Community and the Intelligence Community. Specifically, the program is designed to;

- Reduce the foreign language barriers across the full spectrum of transnational and joint coalition operations

---

14 Department of Defense. *"Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration Community Assistance Response Exercise (CARE) 2004 Assessment Execution Document (AED)"*. May 2004, 2.

- Extend and improve translation capabilities in the coalition military domain

- Expedite access to foreign sources and accelerate processing of foreign language material

- Integrate translation and other language processing tools into IC activities

- Develop tools to improve language learning and sustainment of language skills[15]

Since its inception, the LASER ACTD has included approximately 13 automated language translation tools to allow coalition forces to communicate in multiple languages in real or near real time and to expedite analysis of foreign language or multi-language material. The tools developed through the LASER ACTD were selected to improve coalition task force operations and to improve relations with coalition partners by making them more active participants. The tools also increase the productivity of translators and analysts; enable non-language proficient analysts to take over more of the tasks; and prioritize material for translation and analysis.[16] Many of these tools have been formally and informally evaluated and demonstrated at several international coalition military exercises as well as in local disaster relief exercises and user conferences.

---

15 Office of the Secretary of Defense, *Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD) Management Plan,* November 2003, 8.

16 Ibid., 4.

V.   SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This thesis has attempted to meet its goal of serving one operational purpose and one academic purpose. The operational purpose of providing CONOPS for two specific automated language (human language) translation technology devices has been served in the creation of appendices A and B. Appendix A provides CONOPS for the P2 Phraselator (P2) device and Appendix B provides CONOPS for the Voice Response Translator (VRT) device. These CONOPs will be deployed with the LASER ACTD in the DOD's ongoing effort to pursue ALT technology.

From the academic standpoint, this thesis has attempted to provide the reader with the terminology and framework for understanding the nature and state of current ALT devices. The terminology offered three primary and six secondary descriptors that serve to categorize and compare current ALT devices. Two tables of sample technologies using these descriptors were provided to illustrate these definitions. The notion that human language translation can be accomplished by technology and machines is an appealing one. The notional "Universal Translator" does not exist but there are multiple different devices representing different approaches and methods.

In addition to the terminology and characterization framework, an effort was made to make the reader aware that current ALT devices are still limited but if their limits are understood and trained for, they could be useful in some situations. The human element of utilizing ALT technology possesses certain unique challenges, especially in face-to-face situations. These challenges include expectation, social acceptability or comfort, and socio-cultural differences. For these reasons, the use of an ALT device in a face-to-face situation with a foreign national subject is more subtly difficult than one would expect.

On a national scale, there are tremendous political and military issues associated with human language translation. Both the DOD and the IC need human language processing capabilities in a wide range of languages—for use with both speech and text—to support coalition/joint task force headquarters and tactical or routine field

operations. ALT's can and should increasingly fill this gap, especially as the technologies become more capable.

The potential scope for follow-on study of ALT devices is unlimited but falls roughly into three areas. First, there is room for further study in how to build more effective human training for perspective ALT device users, particularly in face-to-face interactions using speech-to-speech devices. Second, there is a need for further study of the employment of specific devices that take into account the particulars of their limitations, i.e., development of more CONOPS for other devices. Finally, there is a need for constructing a system by which to measure performance of ALT devices.

APPENDIX A.   PROPOSED CONOPS FOR THE P2 PHRASELATOR

CONCEPT OF OPERATIONS
For Conduct of the
P2 Phraselator

Under the Language and Speech Exploitation Resources (LASER) Advanced
Concept Technology Demonstration (ACTD)

TABLE OF CONTENTS

2

1. Purpose: This Concept of Operations (CONOPS) describes the employment of the P2 Phraselator automated language translation (ALT) device. This CONOPS has been developed for the Department of Defense (DOD) Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD). This document is primarily intended for use by the LASER ACTD management team and participating contractors, however, it may be used by other DOD organizations when applicable.

1.1 Background. The generic Phraselator concept was originally developed under a Defense Advanced Research Projects Agency (DARPA) Small Business Innovative Research (SBIR) grant. The need for linguistic services to assist the U.S. military in Afghanistan after September 11, 2001, accelerated the product's development. Shortly after, Phraselator Model 1100 prototypes (the predecessor to the P2) were delivered to US military forces in support of Operation Enduring Freedom (OEF). Since then, continued research has resulted in a new generation Phraselator, called the P2, which is the focus of this document. The P2 Phraselator is a speech-to-speech, one-way translation, phrase-based ALT.

"Speech-to-speech" is translation that is initiated by a voice speaking in the source language into a microphone input and the resulting target language translation is produced audibly via an audio device such as a speaker.

*"One-way translation" is translation only from a source language into a target language. Replies in the target language are **not** translated back. It is imperative that the P2 Phraselator device user have prior training in how to verbally say and understand "yes" or "no" in the target language without the ALT device. Additionally, the user needs to know basic body language gestures of the target culture since these may have different meanings than in American culture. For instance in Iraqi culture, the visual gesture for "no" is one upward nod of the head. This would appear to most Americans to look like "yes" or "go away" and if not understood properly could completely negate any positive effect of operating the ALT device correctly.*

"Phrase-based" translation relies on speech recognition software to identify specific speech input in the source language and match it to a pre-recorded phrase in a target language.

1.2 References.

- U.S. Marine Forces Pacific. *"Demonstration and Assessment Report for Execise Ulchi Focus Lens 2004 Language Translation Systems Limited User Evaluation."* August 2004.
- Department of Defense. *"Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration Community Assistance Response Exercise (CARE) 2004 Assessment Execution Document (AED)"*. FOUO. May 2004.
- Office of the Secretary of Defense. *Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD) Management Plan.* November 2003.

3

27

1.1 Scope:

1.3.1 What It Is. The potential scope of use for the P2 Phraselator is dictated by its capabilities. Since the P2 is a speech-to-speech, one-way, human language translation device that uses strictly pre-recorded phrases, it lends itself best to straightforward and repetitive situations. Any expected replies can be visually expressed by body gestures, compliant behavior, or writing something down on paper. This CONOPS will illustrate the use of the P2 in three environment scenarios; a coalition compound checkpoint, a disaster relief scenario, and a maritime warning operation. This CONOPS acknowledges that there may be other scenarios that can be recorded, rehearsed and utilized but the three depicted scenarios will suffice to illustrate the bulk of its use in a DOD environment.

1.3.2 What It Is Not. The P2 is not a notional "Universal Translator" – meaning it is not a real time, two-way, free-flowing translator – such a device is not technologically feasible yet. The P2 has limitations that require the human user to understand and train for. The biggest challenges for the user are likely to be memorizing and practicing phrase scenarios, practicing quick navigation of the phrase banks in the device, and learning in advance the appropriate human body language gestures of the likely foreign national audience. Additionally, it takes personal poise and human interpersonal skills to stand face-to-face and maintain eye contact with a foreign national subject and read his body language – especially as the foreign national comes to realize it is a machine device talking him.

2.0 Overview.

2.1 Current Situation. On a national scale, there are tremendous political and military issues associated with human language translation. Both the Department of Defense (DOD) and the Intelligence Communities (IC) need human language processing capabilities in a wide range of languages—for use with both speech and text—to support coalition/joint task force headquarters and tactical or routine field operations. Whether handling tactical intelligence or handling foreign national personnel seeking coalition medical assistance, the need for human language translation exceeds the availability of linguists. (LASER MP pg 3) Automated Language Translation Technologies (ALT's) can and should increasingly fill this gap, especially as the technologies become more capable.

2.2 System Summary. There are three physical configurations for use of the P2 Phraselator
    (1) The Basic Configuration (hand held)
    (2) The Megaphone Configuration
    (2) The Long Range Acoustic Device (LRAD) configuration.

2.2.1 Basic Configuration. In the basic configuration, the P2 unit is simply held by an individual person in their hands (figure 2). Additionally, VOXTEC has released a new handsfree version (figure 3)

4

28

Figure 1: The P2 Phraselator

Figure 2: The Basic Configuration

Figure 3: The New Hands-Free Configuration

5

2.2.2 Megaphone Configuration. The P2 Phraselator can be attached to any megaphone to project over longer distances. In this configuration, the user still holds and operates the P2 Phraselator while the megaphone is held in one hand (figure 4). VOXTEC recommends the use of the Minivox megaphone for its durability.

Figure 4: P2 Phraselator Connected to a Minivox Megaphone

2.2.3 Long Range Acoustic Device (LRAD) Configuration. The P2 can be attached to the LRAD to project translated phrases over large distances (figure 5). The P2 Phraselator is connected to the LRAD through either the LRAD MP3 Player or through the MP3 Input connection input directly on the LRAD.

Figure 5: P2 Phraselator connected to an LRAD

6

30

Figure 6: An LRAD

3.0 CONOPS. The P2 Phraselator is a handheld, speech-to-speech, one-way, phrase-based language translation device (figure 1). It takes an input phrase by pushing a Push-to-Talk (PTT) button and speaking into the microphone on top of the device or via the touch screen with a stylus, matching the input with its corresponding translated phrase, and plays that phrase (in the selected target language) through a built-in speaker. The phrases are designed to prompt responses that can be conveyed using gestures such as nodding one's head, holding up a number of fingers, pointing to something, or writing something down on paper. The P2 Phraselator is organized by "Phrase Modules" consisting of groups of phrases and their translations into one or more target languages that represent a specific mission area, such as force protection or medical screening (figure 7). The modules are further divided into subsections for more specific missions such as crowd control or law enforcement (figure 8). The user has the option to create a personal folder and add their most often used phrases to it. This is significant since most of the modules contain hundreds of phrases and it is awkward in face-to-face situations to be searching for more than a few seconds for the next phrase. Due to limited memorization capability, people would naturally gravitate toward a smaller number of immediately available phrases that would work best for them individually.
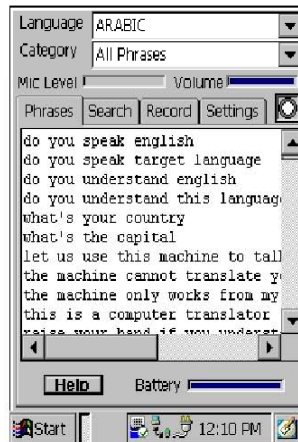

Figure 7: Screen view of P2 module

Figure 8: Screen view of subsections of a phrase module

7

31

Additionally, the P2 Phraselator is often most efficient with two P2 Phraselator familiar people working together. One person, the "user", would hold and operate the device while another team member would render a variety of assistance. The team member's job is to do everything possible to allow the user to smoothly operate the device and maintain control of the situation. The degree of complexity of the situation would determine how often the team member is needed and what he would be doing. For instance in a face-to-face checkpoint scenario, the team member might be needed to search the foreign national subject(s) after the user alerts the subject that he is about to be searched. This allows the user to continue to hold the device, remove his eyes from the subject to look at the screen and scroll as necessary through the phrase list to get to the next appropriate phrase. In situations where there is little face-to-face contact with the subject, such as broadcasting over a megaphone from a distance, there is less complexity for the user so a P2 Phraselator - familiar team member is probably not needed.

If the user is providing input via speech, it is important to note that the desired phrase has to be stated exactly in its entirety in order for the device to recognize it. Since some of the phrases are quite lengthy, the touch screen option using the stylus is more likely to be used. As such, the device often requires the user to look at it, thereby removing his eyes from the foreign national subject.

The P2 is envisioned as a squad level tool for force protection and as a department level tool for medical - in which three people are trained and proficient with it. Since successful use of the device is dependent upon high familiarity and frequent use, it will not likely be effective if everyone in the squad or department tries to get qualified. In recent exercises utilizing ALT devices, it was observed that a few highly adaptable people naturally emerge as the de-facto "experts" because they develop a curiosity and spend time getting familiar with the phrases. The scenarios depicted in this CONOPS exhibit a reasonable breadth of potential use for the devices but are not intended to restrict development of further use scenarios.

The use of the P2 Phraselator will be illustrated utilizing three scenarios;
    (1) A Coalition Compound Checkpoint/Entrance
    (2) A Disaster Relief Scenario
    (3) A Maritime Warning

3.1 Coalition Compound Checkpoint/Entrance. This scenario is positioned in a foreign country where the coalition forces have built or established a physically enclosed compound – similar to establishments in Iraq or Afghanistan today. Coalition personnel who stand guard at the gate can expect to be approached face-to-face by foreign national subjects who may or may not speak English. The guard is responsible for ensuring that nobody enters the compound who is not authorized to and that the subjects are searched for weapons. Depending on the threat situation of the host country, there may be additional security concerns related to insurgency activity and the guards may seek to find out information from potential informants. In the following checkpoint scenario, one of several guards at a checkpoint is holding the P2 Phraselator device and has a team member standing next to him. Both the device user and the team member are familiar and trained on the use of the P2 Phraselator and have constructed a suitable personal folder of their most used phrases respective to checkpoint activities. Both the user and the team member know how to say and understand the words "yes' and "no" in the local language and know the body language gestures associated with "yes" and "no" and how to beckon someone toward them. There are several additional gate guard team members holding

rifles standing in positions around the gate area. Those guards are observing all activity at the gate. The local threat condition is moderate.

3.1.1 Checkpoint Scenario. Two foreign national male subjects in civilian attire approach a coalition compound checkpoint on foot. Neither man is carrying anything in their hands or wearing backpacks. They both are, however, wearing loose flowing robes. Both men look apprehensive but intent on trying to communicate something.

The user looks directly at the approaching subjects and motions for them to approach him. Once they are face-to-face, the user lifts the device to a distance of six inches from his mouth, holds down the Push-to-Talk (PTT) button, states **"This is a computer translator"**, releases the PTT button, points at the P2 Phraselator, and observes the subjects' reaction as the P2 Phraselator broadcasts the translation. The user immediately adds a second message via the PTT button **"Raise your hand if you understand"**.

The foreign national subjects respond by staring at the guard and looking at each other in confusion. The guard realizes the subjects may not speak the target language or are simply shocked by the appearance of an American speaking their language through a machine.

The User activates the two introductory phrases again while maintaining eye contact and observing the body language response of the subjects. This time the subjects appear to focus more closely on the broadcast and then begin saying "yes" in their own language and raising their hands to communicate that they understand the device.

The subjects then begin to point in a direction behind them and talk rapidly in the local language.

The User activates the following phrases in rapid succession using the PTT method **"The machine cannot translate your words for me"**, **"The machine only works from my language to yours"**, and **"raise your hand if you understand"**.

The subjects respond by saying yes in their own language and raising their hands.

The user then initiates a phrase asking **"do you have an appointment here?"**

The subjects respond by saying and visually indicating "no".

The user then stops using the PTT method and shifts his eyes to the screen of the P2 Phraselator while the team member keeps his eyes on the subjects. The user scrolls through his phrase list with the stylus and selects the phrase **"do you have information on anti-coalition activity?"** The user verifies the screen readout in English matches what he selected and conveys to the team member what he is asking (so the team member can follow the context of the conversation).

The subjects excitedly acknowledge "yes.

The user initiates the phrase **"show me your identification"**. The user directs his assisting team member to contact headquarters to see if they can send an interpreter to the gate or an escort to take the men into the compound to an interpreter.

The two subjects offer their ID cards, which the team member takes with him into the guard house to call headquarters.

The user decides he is comfortable taking his eyes off the subject while the team member is in the gatehouse and searches his personal user folder until he finds the following **"would you be willing to make a statement for me to record here?"** and points at the P2 Phraselator.

The subjects indicate "no" they do not want to make a statement.

The user activates the phrase **"describe it with gestures"**.

9

The subjects look confused and make an assortment of unrecognizable gestures with their hands.

The user indicates he does not understand and initiates the phrase **"please wait here"**.

The team member returns from the guard house and indicates headquarters has an interpreter but they want the men brought in. They are sending an escort to the gate. The team member says he has logged in the subjects' ID cards and hands them back to the subjects.

The user initiates the command to the subjects **"You will be escorted inside shortly"** followed by **"I must search you"** and **"are you carrying any weapons?"**

The subjects indicate no, they are not carrying weapons.

The user directs the team member to search the subjects. Upon completion of the search, the user initiates the phrases **"thank-you for your cooperation"** and **"please wait here"**.

3.2 Disaster Relief. This scenario is positioned in an area where a natural disaster has occurred and humanitarian workers are trying to communicate with the local population to render assistance. In the broad scope, relief workers may be performing damaged site assessment and reconstruction, evacuation, missing persons, search and rescue, general distribution of clothing and food, water treatment, sanitation, and medical triage. Some disaster relief scenarios would likely require the use of the P2 Phraselator in both the basic configuration and with a megaphone. In the following specific scenario, which is only one small portion of the possible venues, a team of about 50 relief workers have established a field refugee-type site where the locals are arriving to seek food, water, and medical care. There are several P2 Phraselator teams, each consisting of two people who are both fully trained on the device and have set up their personal user folders with a highly familiar and rehearsed number of phrase particular to their portion of the mission. Two of the teams each set up at separate tables along with other support relief workers, one table for medical and one for other needs. A third team moves up and down the lines of refugees to quickly triage for medical emergencies and make announcements to direct people which line to get into and describe what assistance is available.

3.2.1 Disaster Relief Scenario/Crowd Organization. The roving P2 Phraselator team notes that there appear to be over 100 refugees in the lines approaching the front of the relief station.

The user connects the P2 Phraselator to the megaphone, hands the megaphone to the team member, and then scrolls through the screen display to activate the following announcements: **"we are relief workers here to help"**, **"if you have a medical emergency, please approach me now"**, **"if you are seeking food and water, please join the line on the left"**, and **"if you are seeking non-emergency medical assistance, please join the line on the right"**.

An obviously distraught woman approaches the user and begins speaking in her native language.

The user and the team member note that the woman is very unkempt but has no obvious injuries. The team member makes calming gestures toward the woman while the user disconnects the P2 Phraselator from the megaphone and scrolls through a phrase list. Utilizing the stylus, he activates the phrase **"do you need medical attention?"**

The woman looks surprised for a second and then replies and signals "no".

The User activates the following phrases in rapid succession using the PTT method **"This is a computer translator"**, points at the P2 Phraselator, and observes the subjects' reaction as the P2 Phraselator broadcasts the translation. **"The machine cannot translate your words for**

10

34

me", "The machine only works from my language to yours", and "raise your hand if you understand".

    The woman raises her hand to signal she understands.

    The user then activates the phrase **"do you need water or food?"**

    The woman becomes visibly more upset and starts talking again.

    The user activates the phrase **"are you looking for someone who is missing?"**

    The woman immediately looks relieved and emphatically replies and signals "yes".

    The user and the team member signal for the woman to follow them and they lead her over to an area specially designated for missing person reports

3.3 Maritime Warning. This scenario is positioned in a harbor where small vessels are approaching US Navy ships. This is the most straightforward scenario in that the user does not have close face-to-face contact with foreign national persons. This scenario is <u>not</u> a full blown Maritime Interdiction Operation (MIO) that includes boarding. If it were, the user would have to switch to the Basic Configuration after the vessels were connected and proceed in a face-to-face manner similar to the checkpoint scenario described in paragraph 3.1.1. For this scenario, there is an LRAD with a P2 Phraselator connected to it on the bridge wings of the US Navy ships. Each of the LRAD operators knows how to operate the P2 Phraselator and a has a list of appropriate phrases memorized verbally and collected together in his personal user folder.

3.3.1 Maritime Warning Scenario. A small speedboat of unknown nationality is heading toward a Navy ship.

    The LRAD/P2 Phraselator operator/user broadcasts a pre-recorded warning in English and then initiates a P2 Phraselator command via stylus selection on the screen **"You are approaching a US Navy warship, change your course away from this ship".**

    The user observes the vessel is still continuing inbound, so he then initiates the phrase **"If you do not alter your course, we will fire upon you".**

    The approaching vessel alters its course away from the US Navy Ship

4.0 Logistics.

4.1 P2 Phraselator Maintenance: The P2 Phraselator comes in a pouch containing five components.

    a. Phraselator

    b. Instruction manual; includes User Technical Training instructions.

    c. Instruction mini CD; includes User Technical Training instructions (see section 4.2.2)

    e. Wall outlet charging cord with four detachable plug configurations to accommodate foreign country electrical systems.

    h. Mini USB cable; allows connection to a computer for building phrase files (see section 4.2.3)

4.1.1 P2 Phraselator Maintenance Considerations. It is worth noting that many of the P2 Phraselator components are not specifically marked to be matched with each other. Users at a recent military exercise in Korea frequently misplaced and lost the small pieces. Inventory and accountability are likely to be challenging.

11

Figure 9: The P2 Phraselator bag components and accessories

4.2 P2 Phraselator Training. There are ideally three phases to P2 Phraselator Training.
(1) User Technical Training
(2) User Operational Familiarity Training
(3) Mission Phrase File Build-Up Training

4.2.1 Phase One: User Technical Training. This training refers to the physical set-up of the device where the user learns the components, switches and software features. He learns how to scroll through the visual display screens, and selects a phrase to use either by verbally entering it or by selecting it on the screen with a stylus. He learns how to control the volume and activate other user options such as building his own "favorites" list or configuring the device for left-handed use.
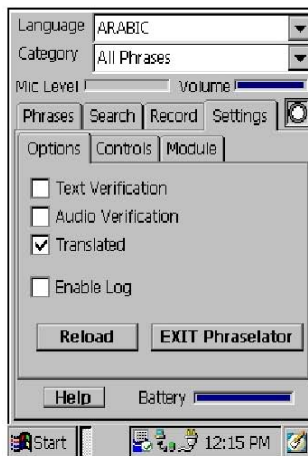

Figure 10: Learning the Options Function

Figure 11: Learning the Record Function

12

36

4.2.2 Phase Two: User Operational Familiarity Training. This is the part of training that is most difficult to learn and is the least appreciated because users tend to "freeze" if they have not rehearsed or gained enough familiarity with the P2 Phraselator to use it effectively while standing face-to-face with a foreign national subject. During Exercise Ulchi Focus Lens 04 in Korea, it was clear that US Marines using the device to communicate with Korean service members were quickly overwhelmed. Although they had completed the User Technical Training described in section 4.2.1 above, the reality of standing face-to-face with a non-English speaking Korean national subject was intimidating and somewhat flustering. This underscores a significant need for high proficiency and familiarity with the device. The US Marines who participated felt that they could do much better with a lot of practice in similar live scenarios. The Marines also asserted they would have to use it frequently to be comfortable with it and to stay proficient with a large number of phrases. This particular terminology, "Phase Two User Familiarity Training", is not formally recognized separately from the Phase One User Technical Training by the industry, although it is generally acknowledged by those who have seen someone try to use the device in a face-to-face situation with a foreign national subject.

      User Operational Familiarity Training includes role playing by the user with foreign national subject actors or linguists. The user has to memorize and gain familiarity with the voice commands and associated translated phrases for predicted scenarios and the user needs to learn basic body language gestures of the anticipated foreign audience. This includes at least how to say and signal "yes" or "no" and how to beckon a person toward them. The user is then placed into a scenario with a foreign national subject actor (or linguist) and has to meet certain performance parameters in his task.

      Because this phase of training is considered so critical, the next section offers a generic set-up for a basic training environment to conduct User Operational Familiarity Training. This proposed training scenario is not set up in a formatted lesson guide in order to facilitate ease of reading within the context of CONOPS. What it should do is offer the reader a fairly specific layout for practice training while not "spoon feeding" the actual phrases. Overall, it offers insight into the scope and necessity of this particular phase of training.

4.2.2.1 Sample Voice Recognition Translation Training Scenario For A Main Gate Sentry Application

<center>ASSUMPTIONS</center>

1. Guard on duty at the gate to a compound understands "yes" and "no" verbally in local language as well as how to gesture for someone to approach.
2. Guard has an assistant to search, verify identification and verify appointment, etc.
3. The foreign speaker speaks a known language.
4. The foreign speaking visitor is a local national subject and is applying for a pass to attend a possibly scheduled meeting with a specific person.

<center>13</center>

ALL SITUATIONS

1. Guard identifies himself and states greeting. Explains about the device he is using (P2 Phraselator) and asks if the visitor can understand what is being said and asks to verify yes by proper body language or to say yes in his language.
2. Guard asks for picture I.D. and do you have an appointment? Yes – No – Visitor gives I.D. to assistant.
3. Assistant verifies I.D. and checks the appointment against a list. If there is an appointment scheduled, the Assistant calls for an escort. If there is no appointment scheduled, the Assistant informs the Guard.
The next six steps only occur if the Guard has determined he will allow the subject to enter the compound.
4. Guard asks, Do you have any weapons? Please answer yes or no in your language.
5. Guard states, If you have any weapons, please surrender them and they will be returned to you when you leave.
6. Guard asks, May we inspect your carry bag and person? Guard directs Assistant to search the subject.

SITUATION #1

The visitor has the proper photo identification, a listed appointment with a known person and no weapons. Utilize the ALL SITUATIONS format above through step 6.

7. Guard states, Your I.D. is acceptable and someone will come to accompany you soon. Please wait for a few minutes. Have you understood? Please say yes or no.

SITUATION #2

Visitor does not have the proper I.D. but has an appointment. Utilize the ALL SITUATIONS format above through step 3.

4. Guard states, Your I.D. is not acceptable. Please obtain the correct I.D. Thank you for your understanding. Good-bye.

SITUATION #3

The visitor has a picture I.D., has an appointment, and has a weapon. Utilize the ALL SITUATIONS format above through step 7.

7. Guard states weapon or contraband cannot pass the gate and must be surrendered. States property will be returned when the visitor leaves.
8. Guard states, Your I.D. is acceptable and someone will come to accompany you soon. Please wait for a few minutes. Have you understood? Please say yes or no

14

Visitor has proper I.D. but does not have an appointment. He is looking for employment. Utilize the ALL SITUATIONS format above through step 3.

2. Guard states, Your I.D. is acceptable, but you do not have an appointment. Please wait and we will contact someone who speaks your language to assist you. Have you understood? Please say yes or no.

MEASURES OF EFFECTIVENESS (MOE's).

These are to be used as a checklist to debrief the user and the team member after each situation is performed.

1. Was the subject's photo ID card checked?
2. Was the subject asked his business such as an appointment or seeking medical help, etc?
3. If the subject indicated he had an appointment, was his ID card checked against an appointment list for verification?
4. If it was determined the subject had a legitimate reason to be admitted, was an appropriate escort called for?
5. Was he/she asked to surrender any weapons?
6. Was the subject then searched?
7. If any weapons were found, were they confiscated and was the subject informed he could collect them upon his departure?

4.2.3 Phase Three: Mission Phrase Group Composition Training. This is the third component of P2 Phraselator training. It is specifically for users and their leadership to identify, learn and build (if needed) specific phrases they need for their missions. Although VOXTEC has already created many groups of potentially useful phrases categorized as "phrase modules", only the military unit who is going to actually use the device can determine the finer details of what they may need to be able to say. The phrases are contained on Secure Digital (SD) cards that can be easily installed in the P2 Phraselator and removed by the user (figure 12).

This training begins by simply reviewing and selecting from available phrase modules that have already been created by VOXTEC. There are presently 32 phrase modules and they are easily accessible online at www.phraselator.com. If the existing phrase modules appear sufficient, the user downloads any combination of modules and languages either via ActivSynch software with a USB interface directly to the P2 Phraselator or by directly writing to an SD card in an SD card reader (figure 13). Either way, the modules are loaded on an SD card.

15

Figure 12: Loading the SD card into the Phraselator

Figure 13: An SD card reader

It is likely during Phase Two User Familiarity Training (discussed in section 4.2.1), the users may find they need some specific phrases that are not in the unit. If the user and his unit need to add more specific phrases, they have three choices. First, they can simply send a list to VOXTEC, who will create a new module. Second, the user can create a new module on a computer using a headset and Voxtec's software called Toolkit Pro . The users would need their own linguist to input the translations. Third, the user can utilize the new field recording feature of the P2 Phraselator Version 3.0 (just released) which allows an input directly to the P2 Phraselator without using a computer. The user simply uses the stylus to "type" in the desired phrase and the linguist speaks it into the device. For situations where the user has arrived in the field and realizes he really needs just one or two additional phrases right way, he can execute this procedure.

VOXTEC continually works with military units to build and update phrase modules. As of February 2005, there are 32 phrase modules available in varying numbers of 41 languages. For instance, 18 of the phrase modules are available in Arabic for a total memory requirement of 57 MB. Only 8 phrase modules (and not necessarily the same modules as Arabic) are available in Thai for a total memory requirement of 27 MB. VOXTEC provides a spreadsheet denoting which phrase modules are available in which specific languages and how much memory is required on the SD card to accommodate each phrase module/language combination. Assuming the user only needs access to all available modules in one or two languages, there is plenty of room on one SD card to contain them plus leave room for field recording. SD cards are currently available in 1GB and higher capacity at any electronics store.

The biggest challenge for phrase group composition is to make the group as short and effective as possible. The limiting factor is how many phrases the user can reasonably be familiar with. The Secure Digital card capacity will allow hundreds of phrases to be recorded but it is unrealistic to expect a human to remember that many. In less tactical situations, phrase look-ups may be possible but they are awkward, especially in face-to-face situations. Diligent attention to this phase of training can ensure that each phrase is worth the trouble of learning it.

5.0 Conclusion. The P2 Phraselator is a speech-to-speech, one-way, phrase based, human language translation device developed by VOXTEC. It is one of several automated language translation devices being evaluated under the LASER ACTD. It can be configured for individual

16

persons to simply hold or it can be connected to a megaphone or to an LRAD. Because the P2 Phraselator is phrase based, the user is required to become familiar with numerous phrases and where they are located in the file structure in order to use the device effectively. Frequent practice and use are necessary to maintain a comfort level that permits the user to maintain composure in a face-to-face situation with a foreign national person. Training is envisioned as having three distinct components, user technical training, user operational familiarity training and mission phrase group composition training. It is envisioned as a squad level device for force protection and as a department level device for medical screening with three trained users to maximize familiarity and proficiency. By limiting the use of the device to straightforward and repetitive situations where any expected replies can be visually expressed by body gestures or compliant behavior, the user can accomplish the mission without the use of a human translator.

17

Appendix A: Acronyms

ACTD Advanced Concept Technology Demonstration
AED Assessment Execution Document
ALT Automated Language Translation
CONOPS Concept of Operations
DOD Department of Defense
IC Intelligence Communities
LASER Language and Speech Exploitation Resources
LRAD Long Range Acoustic Device
MOE Measures Of Effectiveness
PC Personal Computer
SD Secure Digital

18

APPENDIX B.   PROPOSED CONOPS FOR THE VOICE RESPONSE TRANSLATOR

CONCEPT OF OPERATIONS
For Conduct of the
Voice Response Translator (VRT)

Under the Language and Speech Exploitation Resources (LASER) Advanced
Concept Technology Demonstration (ACTD)

TABLE OF CONTENTS

1.0 Purpose. This document describes the Concept of Operations (CONOPS) for employing the Voice Response Translator (VRT) developed for the Department of Defense (DOD) Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD). This CONOPS is primarily intended for use by the LASER ACTD Management Team and participating contractors, however, it may be used by other DOD organizations when applicable.

1.1 Background. The VRT is an automated language translation (ALT) device developed by Integrated Wave Technologies (IWT) of Freemont, California. It translates human language from a source language (the user's language) to a target language (of a foreign national subject). Earlier generations of the VRT were initially fielded in 1997 in civilian police forces as a means of conducting routine traffic stops and crowd control. Later generations have been deployed in DOD since 2000. The VRT is a speech-to-speech, one-way translation, phrase-based tool.

"Speech-to-speech" *is translation that is initiated by a voice speaking in the source language into a microphone input and the resulting target language translation is produced audibly via an audio device such as a speaker.*

"One-way translation" *is translation only from a source language into a target language. Replies in the target language are **not** translated back. It is imperative that the VRT device user have prior training in how to verbally say and understand "yes" or "no" in the target language without the ALT device. Additionally, the user needs to know basic body language gestures of the target culture since these may have different meanings than in American culture. For instance in Iraqi culture, the visual gesture for "no" is one upward nod of the head. This would appear to most Americans to look like "yes" or "go away" and if not understood properly could completely negate any positive effect of operating the ALT device correctly.*

"Phrase-based" *translation relies on speech recognition software to identify specific speech input in the source language and match it to a pre-recorded phrase in a target language. The input can be the phrase itself or a simple command that represents the intended message. For example, the user would say "Hands" into the device in the source language – the device would react by broadcasting "Put your hands in the air" in the target language.*

1.2 References.

- U.S. Marine Forces Pacific. *"Demonstration and Assessment Report for Execise Ulchi Focus Lens 2004 Language Translation Systems Limited User Evaluation."* August 2004.
- Simmonds, Asuncion and Dee Sheppe. Naval Air Systems Command Orlando Training Systems Division. *"Usability Evaluation of Voice Response Translator. Prepared for: United States Special Operations Command."* 12 August 2004
- U.S. Department of Defense. *"Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration Community Assistance Response Exercise (CARE) 2004 Assessment Execution Document (AED)".* FOUO. May 2004.
- Office of the Secretary of Defense. *Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD) Management Plan.* November 2003.

1.3 Scope:

1.3.1 What It Is. The potential scope of use for the VRT is dictated by its capabilities. Since the VRT is a speech-to-speech, one-way, human language translation device that uses strictly pre-recorded phrases, it lends itself best to straightforward and repetitive situations where any expected replies can be visually expressed by body gestures or compliant behavior. This CONOPS will illustrate the use of the VRT in three environment scenarios; a coalition compound checkpoint, a house search, and a maritime warning operation. This CONOPS acknowledges that there may be other scenarios that can be recorded, rehearsed and utilized but the three depicted scenarios will suffice to illustrate the bulk of its use in a force protection DOD environment.

1.3.2 What It Is Not. The VRT is not a notional "Universal Translator" – meaning it is not a real time, two-way, free-flowing translator – such a device is not technologically feasible yet. The VRT has limitations that require the human user to understand and train for. The biggest challenges for the user are likely to be memorizing and practicing phrase scenarios, practicing use of the same voice tone for ease of voice recognition, and learning in advance the appropriate human body language gestures of the likely foreign national audience. Additionally, it takes personal poise and human interpersonal skills to stand face-to-face and maintain eye contact with a foreign national subject and read his body language – especially as the foreign national comes to realize it is a machine device talking him.

2.0 Overview.

2.1 Current Situation. On a national scale, there are tremendous political and military issues associated with human language translation. Both the DOD and the Intelligence Communities (IC) require human language processing capabilities in a wide range of languages—for use with both speech and text—to support coalition/joint task force headquarters and tactical or routine field operations. Whether handling tactical intelligence or handling foreign national personnel seeking coalition medical assistance, the need for human language translation exceeds the availability of linguists. Automated Language Translation technologies (ALT's) can, and should, increasingly fill this gap, especially as the technologies become more capable.

2.2 System Summary. There are three physical configurations for use of the VRT
      (1) The Basic Configuration (hands-free, eyes free)
      (2) The Megaphone Configuration
      (3) The LRAD configuration.

4

Figure 1: The VRT Translator & Headset

2.2.1 Basic Configuration. In the basic configuration, the VRT unit is mounted on an individual person (figure 2). The user wears the headset device and mounts the translator on his vest or in a front pocket. The translator can be mounted in either a standard ammo pouch (figure 3) or by velcro and/or Alice clips (figure 4). This enables the user to wear the VRT and be completely hands-free and eyes-free.

Note that the VRT headset can also be connected through the Modular Integrated Communications Helmet (MICH) headset used by special operations forces. In that instance, the MICH headset would replace the VRT headset.
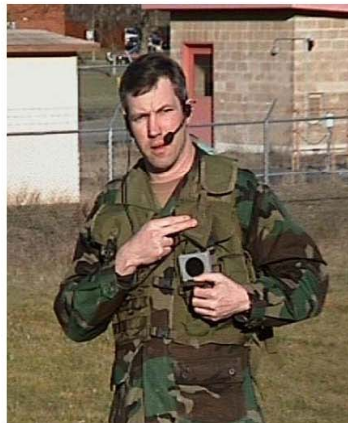
Figure 2: The VRT in the Basic Configuration

5

Figure 3: The VRT speaker in an ammo pouch

Figure 4: The VRT speaker prepared with mounting Velcro and Alice Clips

2.2.2 Megaphone Configuration. The VRT can be attached to the MV-16S Falcon Megaphone to project over longer distances. In this configuration, the user still wears the headset but the VRT translator box is attached to the megaphone (figure 5). The megaphone must be modified to include an input jack for the VRT external speaker cord (figure 6). This modification bypasses the megaphone mouthpiece to ensure there is no acoustic feedback and to provide better overall sound quality. IWT offers Megaphones with the required modifications for users who request it.

Figure 5: The VRT mounted on the MV-16S Megaphone

Figure 6: The modified input jack

6

48

2.2.3 Long Range Acoustic Device (LRAD) Configuration. The VRT can be attached to the LRAD to project translated phrases over longer distances than the megaphone (figure 8). The VRT is connected to the LRAD through either the LRAD MP3 Player or through the MP3 Input connection input directly on the LRAD. If specifically requested by the user, IWT provides appropriate standard audio plugs, e.g. ¼" mono plugs or RCA plugs.

Figure 7: LRAD

Figure 8: VRT attached to an LRAD

3.0 Concept of Operations. The VRT operates by recognizing specific **Voice Commands** from the user and then broadcasting an associated **Translated Phrase**. The voice command must be spoken exactly as it was pre-recorded into the device in order for it to be recognized. For this reason, many of the voice commands are short abbreviations of the translated phrase. For instance, the voice command **"Barricades"** is associated with a translated phrase that says *"Stay behind the barricades"* in the target language. Some sample voice commands and translated phrases are listed below. The composition of phrase lists and where/how they are created is discussed in section 4.2.3.

VOICE COMMAND	TRANSLATED PHRASE
"Begin Directions"	*"I'm speaking to you through a device that translates select phrases into your language. Please respond using hand signals, nodding your head for yes, shaking your head for no, or writing short answers."*
"Barricades"	*"Stay behind the barricades"*
"Turn off engine"	*"Please turn off your engine"*
"Enemy place?"	*"Do you know where enemy soldiers are located?"*
"I say yes"	*"Affirmative"*
"Go this way"	*"Please go this way"*
"Group Leader"	*"Who is your group leader?"*
"Goodbye to you"	*"Good-bye"*

7

49

Because the VRT is phrase based, it requires the user to have memorized the voice commands and the content of its associated translated phrase for a certain number of specific phrases for each mission. The more tactical the mission, the more important it is since there would be no opportunity to search the phrase list. It is estimated that a frequent VRT user could memorize about 50-80 voice commands and their associated translations. This is obviously also a function of individual effort, talent, and how frequently he/she uses the device. Because the VRT is user-dependent, meaning the user has to pre-record his voice to the device, it is necessary for the user to always address the device in the same tone. If the user's voice changes, from stress or other emotion, the device may not recognize the voice command. User familiarity and proficiency could ensure the user is able to stay calm and use the same tone and pronunciation in a challenging situation.

Additionally, the VRT is often most efficient with two VRT familiar people working together. One person, the "user", would wear the device and another team member, would render a variety of assistance. The team member's job is to do everything possible to allow the user to keep his eyes on the subject and maintain control of the situation. The degree of tactical complexity of the situation would determine how often the team member is needed and what he would be doing. For instance in a face-to-face checkpoint scenario, the team member might be needed to search the foreign national subjects or look up an unusual phrase in the phrase-book for the user. In situations where there is little face-to-face contact with the subject, such as broadcasting over a megaphone from a distance, there is less difficulty for the user so a VRT - familiar team member is probably not needed.

The VRT is envisioned as squad level tool, in which three people are trained and proficient with the VRT. Since successful use of the device is dependent upon high familiarity and frequent use, it will not likely be effective if everyone in the squad tries to get qualified. In recent exercises utilizing ALT devices, it was observed that a few highly adaptable people naturally emerge as the de-facto "experts". The scenarios depicted in this CONOPS exhibit a reasonable breadth of potential use for the devices but are not intended to restrict development of further use scenarios.

The use of the VRT will be illustrated utilizing three scenarios;
      (1) A Coalition Compound Checkpoint/Entrance
      (2) A House Search
      (3) A Maritime Warning

3.1 Coalition Compound Checkpoint/Entrance. This scenario is positioned in a foreign country where the coalition forces have built or established a physically enclosed compound – similar to establishments in Iraq or Afghanistan today. Coalition personnel who stand guard at the gate can expect to be approached face-to-face by foreign national subjects who may or may not speak English. The guard is responsible for ensuring that nobody enters the compound who is not authorized to and that the subjects are searched for weapons. Depending on the threat situation of the host country, there may be additional security concerns related to insurgency activity and the guards may seek to find out information from potential informants. In the following checkpoint scenario, one of several guards at a checkpoint is wearing the VRT device and has a team member standing next to him. Both the device user and the team member are familiar and trained on the use of the VRT and have memorized voice commands and the content

8

50

of their associated "translated phrases" suitable to checkpoint activities. The team member has in his possession the laminated quick reference guide for all voice command phrases. Additionally, he has the manual in the gatehouse that includes not only the voice command phrases but also the full translated phrases written out in English in case they need to look up some infrequently used phrases. There are several additional gate guard team members holding rifles standing in positions around the gate area. Those guards are observing all activity at the gate. The local threat condition is high.

3.1.1 Checkpoint Scenario. Two foreign national male subjects in civilian attire approach a coalition compound checkpoint on foot. Neither man is carrying anything in their hands or wearing backpacks. They both are, however, wearing loose flowing robes. Both men look apprehensive but intent on trying to communicate something.

The user looks directly at the approaching subjects, maintaining eye contact and activates an introductory phrase from the VRT by stating "**begin directions**". The VRT device repeats the voice command back to the user in English (to verify it recognized the right input) and then proceeds to broadcast its associated phrase in the target language "*I'm speaking to you through a device that translates select phrases into your language. Please respond if you understand this device by saying "yes" or "no" in your own language*"

The foreign national subjects respond by staring at the guard and looking at each other in confusion. The guard realizes the subjects may not speak the target language or are simply shocked by the appearance of an American speaking their language through a machine.

The User activates the introductory phrase again while maintaining eye contact and observing the body language response of the subjects. This time the subjects appear to focus more closely on the broadcast and then begin saying "yes" in their own language and nodding their heads to communicate that they understand the device.

The subjects then begin to point in a direction behind them and talk rapidly in the local language.

The User initiates the voice command "**need a doctor?**" The VRT repeats the voice command in English so the User is sure it recognized it and then broadcasts the translated phrase "*do you need medical attention?*"

The subjects respond by saying "no" in their language and shaking their heads in a negative manner. They continue to point in a direction behind them.

The User initiates the voice command "**activity info?**" The VRT broadcasts the translated phrase "*Do you have information concerning anti-coalition activity?*"

Both subjects say "yes" in their native language and continue to talk in their language excitedly with emphatic hand gestures and arm waving.

The User is aware that the local population is known for behaving in an animated fashion and calmly directs his team member to contact headquarters for further instruction and a human translator if one is available. He then initiates the voice command ""**Tell how far?**" The VRT broadcasts the translated phrase "*How many kilometers away? Please demonstrate using your fingers.*"

The subjects consult with each other in their language and hold up five fingers.

The user directs his team member to open up a map of the local area and present it to the subjects. He initiates the voice command "**You show me**" and the VRT translates "*show me*".

The subjects point to a specific area on the map and make signals with their hands.

9

The team member informs the user that headquarters has a translator and wants to speak to the men. Headquarters is sending an escort to the gate ASAP.

Since the subjects are going to enter the compound once the escort arrives, the user recognizes they must be searched. He initiates the voice command ""**I must search you**" and the VRT translates "*Before entering the compound, I have to search you.*" He then initiates the voice command "**You have weapons?**" and the VRT translates "*Are you in possession of any weapons?*" The User then says "**Take temporarily**", and the VRT translates "*If so, I must hold onto your weapon while you are in the compound. I will return it to you when you leave.*"

The subjects indicate they do not have weapons.

The user initiates the voice command ""**Escort**" and the VRT translates "*Someone will come soon to escort you*".

The user directs his team member to search the men. Upon completion of the search, the user initiates the command "**wait here**" and the VRT translates "*please wait here*".

3.2 House Search. This scenario is positioned in a foreign country where a small coalition force is searching a neighborhood of homes for weapons caches and insurgent activity. This is a highly tactical scenario with great potential for bodily harm. This scenario is particularly challenging because it requires the use of the VRT in both the megaphone and the basic configurations described in paragraph 2.2 above. One of the Marines in the squad has the VRT device mounted on a megaphone he is holding. He is wearing the headset and has a team member standing next to him. Both the user and the team member are familiar and trained on the use of the VRT and have memorized voice commands and the content of their associated "translated phrases" suitable to house search activities. The team member has in his possession the laminated quick reference guide for all voice commands.

3.2.1 House Search Scenario. A squad of infantry Marines is approaching the first house in a neighborhood attempting to locate insurgency activity. They take positions around the home and hold up the megaphone with the VRT attached.

The user says "**Search for people**" into the VRT headset. The VRT device repeats the voice command back to the user in English (to verify it recognized the right input) and then proceeds to broadcast its associated phrase in the target language "*Warning, United States Marines will be conducting a search of the area in order to look for individuals who are planning attacks against US and coalition forces. We are here to help you. Please be advised that Marines will not hesitate in defending themselves if threatened. We greatly appreciate your cooperation.*"

The user then says ""**House search**" and the VRT translates "*Please open your doors and remain outside in your yard until the search is complete. When the Marines arrive at your house, the homeowner can walk them through the search. We are not here to harm anyone. Our goal is to increase security in the area. Thank you for your cooperation.*"

The door of the house opens and a family of four people exits into the yard.

The user quickly disconnects the VRT from the megaphone and attaches it to his vest. He sets down the megaphone and approaches the head of the family and says "**Begin Directions**". The VRT translates "*I'm speaking to you through a device that translates select phrases into your language. Please respond if you understand this device by saying "yes" or "no" in your own language*"

The homeowner warily says "yes" in his own language.

10

The user says ""**House weapons**" and the VRT translates *"You are permitted to have a weapon to defend your home. The Marines will not seize weapons used for home security if the homeowner identifies them to us before we find them. Should the Marines find unauthorized weapons in the house or yard the homeowner will be apprehended. Please place all authorized weapons outside on the ground, at least ten feet away from any person. Thank you for your cooperation."*

3.3 Maritime Warning. This scenario is positioned in a harbor where small vessels are approaching US Navy ships. This is the most straightforward scenario in that the user does not have close face-to-face contact with foreign national persons. This scenario is not a full blown Maritime Interdiction Operation (MIO) that includes boarding. If it were, the user would have to switch to the Basic Configuration (man mounted) after the vessels were connected and proceed in a face-to-face manner similar to the house search scenario described in paragraph 3.2.1. For this scenario, there is an LRAD with a VRT connected to it on the bridge wings of the US Navy ships. Each of the LRAD operators is wearing the VRT headset.

3.3.1 Maritime Warning Scenario. A small speedboat of unknown nationality is heading toward a Navy ship.

The LRAD/VRT operator/user broadcasts a pre-recorded warning in English and then initiates a VRT voice command **"Stay Away".** The VRT repeats the voice command in English (to verify it recognized the right input) and then broadcasts the associated translated phrase, *"Vessel inbound, vessel inbound, you are approaching a US Navy warship. Alter your course away from this vessel immediately."*

The user then initiates the voice command **"Use deadly".** The VRT broadcasts the translated phrase *"Unidentified vessel, if you fail to stop, deadly force will be utilized".* The user then states **"Fire on You"** and the VRT translates *"I will fire upon your vessel".*

The approaching vessel alters its course away from the US Navy Ship

4.0 Logistics.

4.1 VRT Maintenance: The VRT comes in a pouch containing nine pieces.

    a. Headset.

    b. Translator

    c. Instruction manual; includes User Technical Training instructions as well as the full voice command lists with associated translated phrases.

    d. Set-up CD; includes User Technical Training instructions (see section 4.2.2)

    e. Wall outlet charging cord with four detachable plug configurations to accommodate foreign country electrical systems.

    f. 12 Volt vehicle charging cable; allows charging from a vehicle 12 volt outlet.

    g. BA5590 Charging cable; allows field charging from a BA-5590 battery.

    h. Mini USB cable; allows connection to a computer for building phrase files (see section 4.2.3)

    i. Set of plastic laminate cards that include the voice commands list and a place to write down the user's recorded number.

11

4.1.1 VRT Maintenance Considerations. It is worth noting that many of the VRT components are not specifically marked to be matched with each other. Users at a recent military exercise in Korea frequently misplaced and lost the small pieces. Inventory and accountability are likely to be challenging.

Figure 8: The VRT pouch components and accessories

Figure 9: The basic VRT in its issue pouch

Figure 10: The VRT charger and adaptor pieces

4.2 VRT Training. There are ideally three phases to VRT Training.
  (1) User Technical Training
  (2) User Operational Familiarity Training
  (3) Mission Phrase File Build-Up Training

4.2.1 Phase One: User Technical Training. This training refers to the physical set-up of the device where the user learns the components, switches and knobs. The user then goes through the procedures to pre-record his voice to the device. A recent study commissioned by the United States Special Operations Command (SOCOM) suggests this part of the training can be accomplished in just a couple hours and with minimal instruction beyond the CD or written manual (see references).
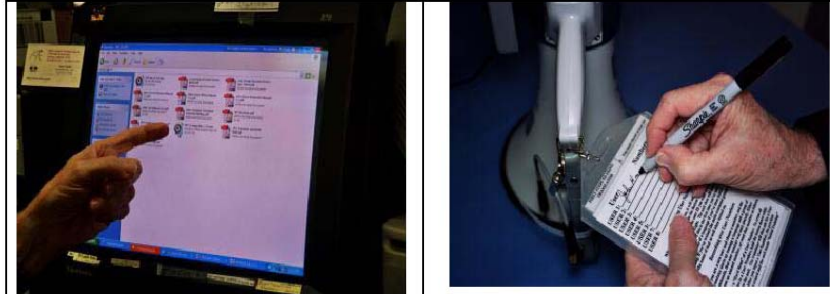
12

Figure 11: The VRT contains a training CD designed to accomplish User Technical training

Figure 12: The user writes down his user # on the laminate quick reference cards.

4.2.2 Phase Two: User Operational Familiarity Training. This is the part of training that is most difficult to learn and is the least appreciated because users tend to "freeze" if they have not rehearsed or gained enough familiarity with the VRT to use it effectively while standing face-to-face with a foreign national subject. During Exercise Ulchi Focus Lens 04 in Korea, it was clear that US Marines using the device to communicate with Korean service members were quickly overwhelmed. Although they had completed the User Technical Training described in section 4.2.1 above, the reality of standing face-to-face with a non-English speaking Korean national subject was intimidating and somewhat flustering. Additionally, several of the users were unable to keep the nervousness out of their voice in the scenarios, to the degree that the device sometimes did not recognize their voice commands. This underscores a significant need for high proficiency and familiarity with the device. The US Marines who participated felt that they could do much better with a lot of practice in similar live scenarios. The Marines also asserted they would have to use it frequently to be comfortable with it and to stay proficient with a large number of phrases.

User Operational Familiarity training includes role playing by the user with foreign national subject actors or linguists. The user has to memorize and gain familiarity with the voice commands and associated translated phrases for predicted scenarios and the user needs to learn basic body language gestures of the anticipated foreign audience. This includes at least how to say and signal "yes" or "no" and how to beckon a person toward them. The user is then placed into a scenario with a foreign national subject actor (or linguist) and has to meet certain performance parameters in his task.

Because this phase of training is considered so critical, the next section offers a generic set-up for a basic training environment to conduct User Operational Familiarity Training. This proposed training scenario is not set up in a formatted lesson guide in order to facilitate ease of reading within the context of CONOPS. What it should do is offer the reader a fairly specific layout for practice training while not "spoon feeding" the actual phrases. Overall, it offers insight into the scope and necessity of this particular phase of training.

13

4.2.2.1 Sample Voice Recognition Translation Training Scenario For A Main Gate Sentry Application

ASSUMPTIONS

1. Guard on duty at the gate to a compound understands "yes" and "no" verbally in local language as well as how to gesture for someone to approach.
2. Guard has an assistant to search, verify identification and verify appointment, etc.
3. The foreign speaker speaks a known language.
4. The foreign speaking visitor is a local national subject and is applying for a pass to attend a possibly scheduled meeting with a specific person.

ALL SITUATIONS

1. Guard identifies himself and states greeting. Explains about the device he is using (VRT) and asks if the visitor can understand what is being said and asks to verify yes by proper body language or to say yes in his language.
2. Guard asks for picture I.D. and do you have an appointment? Yes – No – Visitor gives I.D. to assistant.
3. Assistant verifies I.D. and checks the appointment against a list. If there is an appointment scheduled, the Assistant calls for an escort. If there is no appointment scheduled, the Assistant informs the Guard.
The next six steps only occur if the Guard has determined he will allow the subject to enter the compound.
4. Guard asks, Do you have any weapons? Please answer yes or no in your language.
5. Guard states, If you have any weapons, please surrender them and they will be returned to you when you leave.
6. Guard asks, May we inspect your carry bag and person? Guard directs Assistant to search the subject.

SITUATION #1

The visitor has the proper photo identification, a listed appointment with a known person and no weapons. Utilize the ALL SITUATIONS format above through step 6.

7. Guard states, Your I.D. is acceptable and someone will come to accompany you soon. Please wait for a few minutes. Have you understood? Please say yes or no.

SITUATION #2

Visitor does not have the proper I.D. but has an appointment. Utilize the ALL SITUATIONS format above through step 3.

14

4. Guard states, Your I.D. is not acceptable. Please obtain the correct I.D. Thank you for your understanding. Good-bye.

SITUATION #3

The visitor has a picture I.D., has an appointment, and has a weapon. Utilize the ALL SITUATIONS format above through step 7.

7. Guard states weapon or contraband cannot pass the gate and must be surrendered. States property will be returned when the visitor leaves.
8. Guard states, Your I.D. is acceptable and someone will come to accompany you soon. Please wait for a few minutes. Have you understood? Please say yes or no

SITUATION #4

Visitor has proper I.D. but does not have an appointment. He is looking for employment. Utilize the ALL SITUATIONS format above through step 3.

2. Guard states, Your I.D. is acceptable, but you do not have an appointment. Please wait and we will contact someone who speaks your language to assist you. Have you understood? Please say yes or no.

MEASURES OF EFFECTIVENESS (MOE's).
These are to be used as a checklist to debrief the user and the assistant after each situation is performed.

1. Was the subject's photo ID card checked?
2. Was the subject asked his business such as an appointment or seeking medical help, etc?
3. If the subject indicated he had an appointment, was his ID card checked against an appointment list for verification?
4. If it was determined the subject had a legitimate reason to be admitted, was an appropriate escort called for?
5. Was he/she asked to surrender any weapons?
6. Was the subject then searched?
7. If any weapons were found, were they confiscated and was the subject informed he could collect them upon his departure?

4.2.3 Phase Three: Mission Phrase Group Composition Training. This is the third component of VRT training. It is specifically for users and their leadership to build and learn specific phrases they need for their missions. Although IWT has already created many groups of potentially useful phrases categorized by mission, only the military unit who is going to actually use the device can determine the finer details of what they may need to be able to say.

This training begins by simply reviewing and selecting from available phrase group modules that have already been created by IWT. Assuming the user and his unit need to add

15

57

more specific phrases, they have two options. The first option is to compile their list of additional phrases and forward them to IWT where they will be loaded onto compact flash (CF) cards. The CF cards can be loaded into the VRT unit by the user. IWT continually works with military units to build and update phrase modules.

The second option is for users to load the new phrases into the VRT's by themselves using the VRT application software. The IWT VRT software program, which is used to create VRT applications, is proprietary. Application files including sound files, are initially stored in a directory/folder on a personal computer (PC) with Microsoft Windows Operating System. Then, the IWT program is used to assemble these files into VRT application files. These application files are then transferred to the CF card, which is then loaded into the back of the VRT (figure 13). Procedures are provided by IWT for units who want to download and directly use the VRT application.

Figure 13:The Compact Flash (CF) card being loaded into the VRT

Units may later re-evaluate phrase groups after using them on deployment. It will be likely that new phrases need to be added after arriving in country and experiencing the environment. The VRT incorporates a field recording device that allows a limited amount of new phrases to be added directly to the VRT without utilizing the PC application software and with the assistance of a linguist.

The biggest challenge for phrase group composition is to make the group as short and effective as possible. The limiting factor is how many phrases the user can reasonably be familiar with. The memory chip of the VRT will allow hundreds of phrases to be recorded but it is unrealistic to expect a human to remember that many. In less tactical situations, phrase look-ups may be possible but they are awkward, especially in face-to-face situations. Diligent attention to this phase of training can ensure that each phrase is worth the trouble of learning it.

16

4.2.3.1 Sample Mission Phrase Group. The following list offers a selection of translated phrases that might be needed for the main gate sentry application presented in section 4.2.2.1 above.

VOICE COMMAND	TRANSLATED PHRASE
"My greetings"	*"Good day – I am the Guard at this gate."*
"Begin directions"	*" I am speaking to you through the use of an electronic device that translates a limited number of phrases into other languages. Do you understand what I am saying? Please respond in your language yes or no."*
"Eye-dee"	*"Please show me your picture I.D. card"*
"A meeting?"	*"Do you have a scheduled appointment? Please answer in your language yes or no".*
"Meeting who?"	*"Can you write the name of the person you are meeting with and the time of your appointment?"*
"Do you understand?"	*"Do you understand what I have just said? Please answer yes or no."*
"Any weapons?"	*"Do you have in your possession any weapons? Please answer yes or no"*
"Take temporarily"	*"If you have any weapons, show them to me; you must surrender them. They will be returned to you when you are ready to leave."*
"Any more weapons?"	*"Are these the only weapons you have? Please answer yes or no.*
"Personal search"	*"Please allow my assistant to search your person and bag."*
"I thank you"	*"Thank you"*
"Please wait"	*"Please wait here."*
"Eye-dee is good"	*"Your I.D. is acceptable"*
"You may pass"	*"You may pass"*
"Escort"	*"Someone will come soon to escort you".*

5.0 Conclusion. The VRT is a speech-to-speech, one-way, phrase based, human language translation device developed by Integrated Wave Technologies. It is one of several automated language translation devices being evaluated under the LASER ACTD. It can be configured for individual persons in a hands-free, eyes-free manner or mounted to a megaphone or to an LRAD. Because the VRT is phrase based, the user is required to become familiar with numerous voice command phrases and the content of their associated translated phrases in order to use the device effectively. Frequent practice and use are necessary to maintain a comfort level that permits the user to maintain composure and the same voice tone in the operational environment. Maintaining the same voice tone ensures the user's voice is correctly recognized by the device and contributes to the user's overall control of a face-to-face situation with a foreign national person. Training is envisioned as having three distinct components, user technical training, user operational familiarity training and mission phrase group composition training. It is envisioned

17

as a squad level device with three trained users to maximize familiarity and proficiency. By limiting the use of the device to straightforward and repetitive situations where any expected replies can be visually expressed by body gestures or compliant behavior, the user can accomplish the mission without the use of a human translator.

18

Appendix A: Acronyms

ACTD Advanced Concept Technology Demonstration
AED Assessment Execution Document
ALT Automated Language Translation
CF Compact Flash
CONOPS Concept of Operations
DOD Department of Defense
IC Intelligence Communities
IWT Integrated Wave Technologies, Inc.
LASER Language and Speech Exploitation Resources
LRAD Long Range Acoustic Device
MICH Modular Integrated Communications Helmet
MIO Maritime Interdiction Operation
MOE Measures Of Effectiveness
PC Personal Computer
SOCOM United States Special Operations Command
VRT Voice Response Translator

19

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C: ABBREVIATIONS AND/OR ACRONYMS

ACTD   Advanced Concept Technology Demonstration
AED   Assessment Execution Document
ALT   Automated Language Translation

CF   Compact Flash
CONOPS   Concept of Operations

DOD   Department of Defense

IC   Intelligence Communities
IWT   Integrated Wave Technologies, Inc.

LASER   Language and Speech Exploitation Resources
LMUA   Limited Military Utility Assessment
LRAD   Long Range Acoustic Device
LUE   Limited User Evaluation

MB   megabytes
MIO   Maritime Interdiction Operation
MOE   Measures Of Effectiveness
MT   Machine Translation
MUA   Military Utility Assessment

OCR   Optical Character Recognition

PC   Personal Computer
PDA   Personal Digital Assistant

SD   Secure Digital
SOCOM   United States Special Operations Command

TM   Translation Memory

VRT   Voice Response Translator

THIS PAGE INTENTIONALLY LEFT BLANK

BIBLIOGRAPHY

Air Force Operational Test and Evaluation Center (AFOTEC), Detachment 1. *"Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD) Community Assistance Response Exercise (CARE) 2004 Limited Military Utility Assessment (LMUA) Report."* July 2004.

Air Force Operational Test and Evaluation Center (AFOTEC), Detachment 1. *"Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration Community Assistance Response Exercise (CARE) 2004 Assessment Execution Document (AED)"*. May 2004.

Breault, Chris of the US Marine Forces Pacific Experimentation Center. . Private conversations 13 Aug 04 through 18 Oct 04.

Breault, Chris. *"Language and Speech Exploitation Resources ACTD"*. In *Advanced Concept Technology Demonstration and Transformation Workshop*, United States Pacific Command. Honolulu, Hawaii 8-12 March 2004, CD-ROM.

Director, Marine Corps Forces Pacific Experimentation Center. *"Letter of Instruction (LOI) for Marine Corps Forces Pacific Experimentation Center's Conduct of Language Translation Tool Assessment During Exercise Cobra Gold 04"*. 5 April 2004.

Ehsani, Farzad of Sehda Inc. Private telephone conversation 28 Feb 05. Monterey, CA.

Hall, John of Integrated Wave Technologies. Private telephone conversations 29 Nov 04 through 7 Dec 04. Monterey, CA.

Hall, John of VOXTEC. Private telephone conversations 18 Feb 04 through 7 Mar 04. Monterey, CA

LeBlanc, Ray of MITRE Corporation. Private telephone conversations 28 Feb 05 through 3 Mar 05. Monterey, CA

Nothing Lost in the Translation. *Tech Beat*, Fall 2003. http://www.nlectc.org/techbeat/fall2003/NothlostFall03.pdf, (Accessed 18 Oct 04).

Office of the Secretary of Defense. *Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD) Management Plan.* November 2003.

Phraselator Model P2 web-page. http://www.phraselator.com/products/prod_p2.aspx (Accessed 27 Feb 05)

Sehda Inc. Solutions S-Minds web-page. http://www.sehda.com/solutions.htm. (Accessed 21 Feb 05).

Speechgear Compadre Expres web-page. http://www.speechgear.com/compadre.aspx
(Accessed 25 Oct 04)

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. James F. Ehlert
   Naval Postgraduate School
   Monterey, California

4. Steven J. Iatrou
   Naval Postgraduate School
   Monterey, California

5. Chris Breault
   U.S. Marine Corps Forces Pacific Experimentation Center
   Carlsbad, California

6. CAPT Jeff Kline USN
   Naval Postgraduate School
   Monterey, California

7. Farzad Ehsani
   Sehda Inc.
   Mountain View, California

8. Ann Libengood
   Family Member (sister)
   Boise, Idaho

9. John Nesmith
   Meridian Automotive and Machine Inc.
   Meridian, Idaho